Library of Congress Working Group on the Future of Bibliographic Control

"Economics and Organization of Bibliographic Data" RESPONSE TO BACKGROUND PAPER

Karen Calhoun, OCLC Online Computer Library Center, Inc.

July 9, 2007

ECONOMIC CHALLENGES

1. **Productivity.** Fred Kilgour's genius, when he founded OCLC in 1967, was his coupling of the technological breakthrough of the MARC record with the notion of a shared computerized cataloging system for 54 Ohio college libraries [1]. He recognized and removed the redundancy of rooms of catalogers, in library after library, typing cards for the same books [2]. Over time Kilgour's inventiveness helped libraries achieve double digit (and more) increases in productivity, lowering operational costs, freeing up staff, and providing a new plane on which cooperative endeavors could flourish [3].

   The transformative power of the Internet and the Web is a constant refrain. For all its disruptive impact on libraries, the far-reaching connectivity of the Web has made it possible to take a fresh look at enhancing productivity anew by reducing other duplicative activities across libraries. For example, OCLC's new end-user services Open WorldCat, WorldCat.org and WorldCat Local improve productivity by leveraging libraries' investments in their collections and in the collaboratively-built OCLC database. The results are greater visibility of library holdings and a better discovery service that enables end-users to move more easily back and forth between global, group, and local library collections that are available to them.

2. **Redundancy.** Web connectivity is already being harnessed to help create and share knowledge and works of the imagination that are packaged as books. Most innovations so far have occurred in the private sector, where system design began with *tabula rasa*, without prior tradition or disposition. Contrast this approach with the incremental automation of library acquisitions, whose manual practices were in place decades before anyone thought about library systems, let alone the Web. As a result, acquiring print monographs in libraries today involves redundant, independent connections to publishers and vendors as each library uses its own

ILS to order, pay for, and receive new books. As an example of a different paradigm for library acquisitions, consider how Amazon takes advantage of Web technologies to host many booksellers' wares, or how e-Bay's PayPal connects buyers and sellers on the network. If we were to design library book acquisitions anew, would libraries not also take advantage of the power of the network and more collaborative, interconnected systems?

Redundant effort in cataloging—and its effects, high costs and slow availability of new library materials—has been a persistent issue in our profession. Andrew Osborn, an influential librarian in his day, said in 1936, "real economy can only come through big changes in cataloging methods, through an altered conception of scholarship in relation to library technique and through a willingness to cooperate and to accept the results of cooperation to the full" [4]. Bemoaning catalogers' extensive review and alteration of other catalogers' work (and its justification on the basis of scholarly needs), Osborn noted that "the world of scholarship is oblivious of many of the refinements of cataloging." One might argue that in the ensuing seventy-one years, catalogers have generally become more accepting of one another's work, but technical services managers in large U.S. research libraries continue to express frustration about extensive backlogs and the high cost of cataloging.

3. **Value.** Question 1 of the background paper asks "What trade-offs are being made between quality and economic constraints?" While professionals in our field use the word "quality" often, and we think we know what it means, I believe we are referring to a broad range of meanings—each consistent with a particular set of assumptions.

   OCLC hears many voices speak of quality. When OCLC receives a call from a catalog librarian, and the caller speaks of quality, he or she is most likely referring, in his or her role as an expert, to conformance to bibliographic control standards of some kind. When OCLC speaks with advisory groups of library decision makers, these individuals tend to speak of quality from a variety of perspectives: is the library able to maintain a balance between what comes in for processing and what is completed? Are important new initiatives successfully budgeted and staffed? Are funders satisfied with the value the library brings to the community? Are end users satisfied with the value the library brings to them?

   There is still another voice, the end-user's. LibQual results for 2007 suggest that "quality" for end users in colleges and universities

involves personal control over their access to information. The largest gaps between desired and perceived levels of service in the aggregated 2007 LibQual survey results has to do with (1) library Web sites' enabling users to find information on their own; (2) remote access from their homes and offices; and (3) print and e-journal collections they need for their work [5]. Many other studies are beginning to appear that make the end user's voice on the question of quality more audible to us.

The meanings of "quality" and "value" are thus more complex than we may realize. From an economic perspective, the *value* of bibliographic control reflects not only the costs of creating, copying, exporting, and maintaining records over time, it also reflects the return on the investment, or the value for us today of predicted long-term benefits to the communities we serve, less the costs of producing these benefits [6]. Good stewards of the financial resources invested in libraries will inevitably—and quite correctly—look for balance between long-term benefits and costs, and these librarians will characterize their choices as responsible leadership rather than "trade-offs between quality and economic constraints."

Returning to Osborn's 1936 article, he names one factor affecting cataloging costs as "a belief in many quarters that what Cutter called full cataloging is uniformly desirable, and that what he called medium and short are dangerous playthings" [7]. Similarly, when we characterize simpler catalog records as merely "good enough," we weight the bibliographic control *expert's* definition of quality at the expense of more pragmatic, mission-based, or user-centered perspectives of quality. "Good enough" can actually be better, depending on the circumstances. Further, "full" can actually refer to enrichment through automated means (tables of contents, marketing information, delivery options) or user contributions (reviews, recommendations, tags and other interactive data), rather than the fullness of the bibliographic record according to the cataloging code.

4. **Scale.** Traditional library materials continue to pour into acquisitions and cataloging, while at the same time electronic resources and digital collections demand new workflows, additional metadata types and standards, technology-based processing methods, and new tools. In his talk at the ALCTS 50th anniversary conference last month, David Lankes, Syracuse University School of Information Studies, outlined the issues that libraries face in today's "massive scale" environment [7]. Lankes warns of the

impossibility of cataloging it all using today's methods. At the same time, he also rejects what he calls "bibliofundamentalism" and the option of the library's retreating to the position of a niche player in its community. Instead Lankes favors an approach called "participatory librarianship" [8].

5. **Budgets.** The following section lays out how Association of Research Library (ARL) budgets are allocated based on averages calculated from the 2005 data set [9]. The materials budget (43%) and the allocation to salaries and wages (45%) continue to account for the lion's share, with hardware, software, payments for utilities and networks, document delivery, etc. coming from the remaining 12%. With serials (e- and print) now claiming 67% of the materials budget (with e-resources claiming half of that), most ARL directors are likely to consider the materials budget sacred.

   As a result I have observed that new library initiatives tend to be funded either through whatever discretionary funding can be gathered together or through the salaries and wages budget. With technical services staffing accounting for from twenty to twenty-five percent of the salaries and wages budget in the largest ARL libraries, the intensive budget pressures experienced over the last ten years in technical services departments are not surprising, as library decision makers scramble to fund a variety of new digital initiatives and the rising costs of information technology (both manpower and systems).

   Pressure on technical services staffing budgets is occurring at the same time as ARL materials expenditures continue undiminished [11]. This is an important point, because technical services staffing is declining while the number of materials needing processing is not [12]. Furthermore, the manpower for processing physical resources (books, serials, audiovisual materials) continues to demand a large share of technical services' salaries and wages; ARL libraries continue to acquire tens of thousands of printed books each year; and 85% of WorldCat is still books (e-books account for less than 1% of this number) [13].

   My intuition suggests that some sort of tipping point is coming. The large expenditures on e-resource aggregations (and the necessity of making them accessible to users) have already had a dramatic impact on the practice of serials librarianship, bringing new automated techniques for record creation and maintenance and demanding new job descriptions, skill sets and tools. The enormous

wave of change that passed through serials librarianship in the last ten years has not yet reached monographs cataloging in any significant way, but I believe it will.

6. **Demography.** For years Stanley Wilder, University of Rochester River Campus Libraries, has been studying the demographics and hiring patterns of ARL librarians. Based on his analysis of ARL salary survey and Current Population Survey data, he reported that as a group, North American library professionals were older than comparable professionals [14]. In 2002, based on his continued study of demographic trends and hiring practices, he reported that catalogers were, as a group, even older [15]. In the fifteen years leading up to the year 2000, while the hiring of "functional specialists" (new types of jobs in libraries, often technology-based) rose dramatically, new hires to cataloging positions fell 45%. During that same fifteen years, the hiring of newly-degreed librarians to cataloging positions fell 64%.

   Where are the professional cataloging jobs going? Overall, library staffing has remained constant. In his latest publications, Wilder further describes the rise of a new kind of library professional, with new skill sets, who is more like to work in systems, human resources, fund raising, another other nontraditional jobs [16, 17, 18].

   The retirement wave for a generation of bibliographic control experts is expected to build to full strength starting in 2010. After the exodus that is coming, it seems to me unlikely that the role of librarians in technical services departments will continue in the same way, considering the competing pressures on the salaries and wages budget and new hiring patterns described previously. I fear that research libraries will be unable to sustain the traditional practices and staffing patterns of bibliographic control, whether they wish to or not.

7. **Collaboration.** The economic challenges discussed here will not be met by individual libraries working alone. As our profession defines what bibliographic control is to become in a Web world, I offer some suggestions of how OCLC might help libraries boost their productivity to new levels. Lest the reader misconstrue my purpose, I suggest these technological, people- and process-intensive changes not for the sake of cost reductions as ends in themselves, but to free up human and financial resources that libraries can redeploy on behalf of the communities they serve.

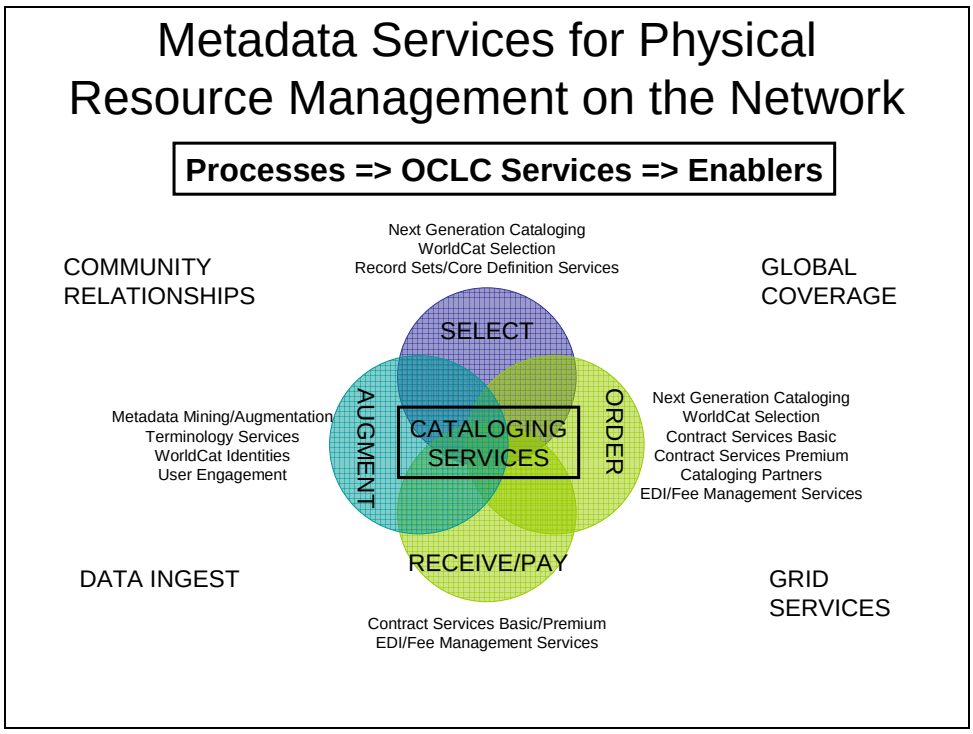
OCLC AND THE ORGANIZATION OF COMMUNITY METADATA SERVICES

OCLC provides both consumer and library management services. Of late OCLC has invested in new consumer services to make local, group, and global library collections more visible and dramatically improve users' experiences. OCLC's library management services have improved incrementally over the past ten years, and the tools that OCLC provides to support bibliographic control remain highly successful. However, the "massive scale" environment that I referred to earlier—or what we at OCLC have begun to call "Web scale"—poses challenges to library operations and workflow that we need to work together to meet.

The following section provides some ideas about how OCLC might help libraries by more actively partnering in the publication supply chain and by augmenting the present cataloging service. The strategy rests upon the goal of further leveraging the world's largest aggregation of bibliographic data, WorldCat. The purpose is to help libraries re-engineer their workflows to lower the costs of acquisitions and cataloging while improving the usefulness of records. I hasten to add that with a few exceptions (for example, WorldCat Selection) the ideas are just that—ideas—not actual or even planned services. Also, for a variety of reasons, I have limited the scope today to physical resource management (not e-resource or digital collections management).

The figure [19]—an integrated framework of sorts—presents library internal processes in the center of the diagram as *one* interdependent flow (rather than multiple independent flows), starting with selection and proceeding through ordering, receiving and payment, cataloging, and catalog maintenance. Further, the framework assumes a Web 2.0 world [20], in which coverage is global; the network is the platform; and data sources are re-mixed, re-used, and syndicated. Last, the framework assumes that metadata (descriptive, administrative, evaluative) has a life cycle. Thus the notion of a "dynamic" metadata record, which develops and becomes fuller over time, is a key element of the framework.

In the circle outside the processes are potential or existing OCLC services to help libraries achieve efficiencies at each point in the flow. Note, cataloging can happen at any point during the flow. At the outermost boundary of the figure are a number of enabling services, explained below.



1. Enablers

- Global coverage. OCLC has moved vigorously in the last few years to legitimize the use of the term “world” in WorldCat. In the last year, OCLC has loaded millions of non-U.S. records and expects to add many more in the next year. Last year’s annual report noted that about 43% of the records in WorldCat are for materials in languages other than English; this trend continues to strengthen [21].

- Grid services. In the same way that the North American power grid interconnects separate electric generating, transmitting, and distribution systems, library grid services are intended to facilitate partnerships, syndication and data exchange, for example to underpin the creation of a virtual publications supply chain connecting participating publishers, vendors, libraries, and fulfillment services.

- Data ingest. The realization of a next generation of OCLC library services will require much better support at OCLC for batchloading and other types of data ingest, conversion, mapping, extraction, and data transfers.

- Community relationships. Libraries and OCLC have a long history of collaboration, but achieving dramatic productivity improvements in acquisitions and bibliographic control will require OCLC to engage in an unprecedented level of collaboration with, between, and outside libraries.

2. Selection and Acquisitions

The concept of "next generation cataloging" is to move bibliographic data capture and initial augmentation upstream to the source of the supply chain, through partnerships with participating publishers and vendors. WorldCat Selection, introduced by OCLC a few months ago, pulls a variety of vendor offerings into one interface and produces bibliographic and ordering metadata as a by-product of selection. The concept of "record sets/core definition services" is to allow selectors to identify a core set of materials (say, a popular reading collection for a Korean community residing in a large U.S. urban center), select them, and receive both the materials and catalog records together.

The concept of "contract services basic" is to provide a simple, technology-based, inexpensive contract cataloging service that libraries can use, for example, to eliminate backlogs. OCLC Contract Services—which I have termed "premium"—already exists, as does the Cataloging Partners program [22]. The idea is to offer libraries many options for capturing bibliographic records at various levels and stages in the workflow. The concept of "EDI/fee management services" is still very soft, but it would involve exploring whether it is feasible to build on the idea of ILL Fee Management [23] to support library selection and acquisitions.

3. Metadata Mining and Augmentation Services

WorldCat is an incredibly rich but largely untapped resource for creating linkages between classification and subject terminology. WorldCat data mining could support the automated augmentation of bibliographic records or cataloger desktop productivity tools. The concept "metadata mining and augmentation" would start from the valuable work already done by the OCLC Office of Research—for example, the Terminologies Service and WorldCat Identities [24]—and extend this work to create new efficiencies in acquisitions and bibliographic control.

4. User Engagement

In his article on Web 2.0, Tim O'Reilly, speaking of Amazon's success, wrote "Amazon has made a science of user engagement. They have an order of magnitude more user reviews, invitations to participate in varied ways on virtually every page—and even more importantly, they use user activity to produce better search results" [25]. In its user-facing service, WorldCat.org, OCLC has also invited user participation in metadata enrichment and personalization services, the latest being a list-making tool [26]. The concept of "user engagement" in the context of this paper is to not to rely on catalogers alone to augment records, but to feed publisher- or vendor-supplied data, user-supplied metadata, and other externally-supplied data (say from a developers network) back into acquisitions and bibliographic control workflows—thereby making *their* data work harder.

A WORD ON POINT OF VIEW

This paper reflects a perspective heavily influenced by my work experience in the member libraries of the Association for Research Libraries (ARL). To counter this bias, I have engaged some of my public library colleagues in discussion about the background paper for this meeting and, with the help of the staff of the OCLC Information Center, gathered a few statistics from individual public libraries (that are not research libraries). I haven't learned enough to counter my bias, but I have learned enough to realize that a public library perspective on the future of bibliographic control is likely to be quite different from my own. Based on my understanding of the issues so far, this is due to significant differences between the way academic research and public libraries allocate their budgets, especially the materials and manpower budgets; the nature of the materials collected; the amount of cataloging copy that is available, and when; the kinds and numbers of unique materials that need to be organized; and more. Two suggestions from my colleague George Needham, Vice President of OCLC Member Services, former State Librarian of Michigan, and former Executive Director of the Public Library Association, were to learn more about the public library perspective on bibliographic control by engaging with the Urban Libraries Council and exploring the data set upon which the Public Library Data Service bases its reports.

NOTES AND REFERENCES

[1] Kilgour, F.G. 1984. Initial system design for the Ohio College Library Center: a case history (p. 103-9). In *Collected papers of Frederick G. Kilgour*, edited by Lois L. Yoakam. Dublin, OH: OCLC. First published in *Proceedings of the 1968 clinic on library applications of data processing* (p. 79-88). Urbana, IL: University of Illinois, Graduate School of Library Science, 1969.

[2] Not all records were done from scratch of course. A useful history of the LC card program can be found in Tillett, Barbara B. 1993. Catalog it once for all: a history of cooperative cataloging in the United States before 1967. *Cataloging and classification quarterly* 17:3/4, p. 3-38.

[3] For a more detailed treatment of the issue of productivity in libraries, see Calhoun, Karen. 2003. Technology, productivity and change in library technical services. *Library collections, acquisitions, and technical services* 27, p. 281-289.

[4] Osborn, Andrew D. 1936. Cataloging costs and a changing conception of cataloging, p. 50, 51. In *Catalogers' and classifiers' yearbook 1929-1945*, p. 45-54. Chicago, IL: American Library Association, Division of Cataloging and Classification.

[5] 2007 LibQUAL Survey Highlights: http://www.libqual.org/documents/admin/LibQUALHighlights2007_SessionI.pdf

[6] A long-term benefit in this context may or may not be monetary. A non-monetary example might be the number of users who use a bibliographic record for some purpose, whether that use is inventory control, discovery, citation, or circulation of library materials.

Some may argue that the notions of "return on investment" and economic value are irrelevant in libraries. Are they? It is true that scholarship, intellectual freedom, the contribution of libraries to research, teaching and learning—and yes, bibliographic control—are properly understood as socially desirable goods that are and should be subsidized and protected from market forces. Nevertheless, individuals and governments do make choices about socially desirable goods, because the amount of resources available to society is always fixed, and choices must be made. For example, the resources that are

invested in interstate highways are not available for investment in alternatives such as large scale, modern public transportation. So, competition and economic forces are active determinants in both the marketplace and public sector. The resources we as a society collectively invest in scholarship, public higher education and libraries are not available for investment elsewhere. And, the resources that are invested in library catalogs are not available for investment elsewhere in libraries.

[7] Osborn, p. 48.

[8] Lankes, R. David. "Collecting conversations in a massive scale world." Presentation at the ALCTS 50th Anniversary Conference, June 21, 2007, Washington DC.
http://quartz.syr.edu/rdlankes/Presentations/2007/ALCTS.pdf

[9] Lankes, R. David. "Reference: an island of chaos in a sea of order." Presentation at the Pratt School of Information and Library Science, April 30, 2006, New York NY.
http://quartz.syr.edu/rdlankes/Presentations/2006/RefChaos.pdf

[10] Association of Research Libraries. ARL Statistics Interactive Edition. Summary Statistics. http://fisher.lib.virginia.edu/cgi-local/arlbin/arl.cgi?task=setupstats

[11] The median of ARL library materials expenditures has increased an average of 6.4% a year since 1986 (unadjusted dollars). Serials expenditures have risen more sharply (7.6%), while monograph expenditures have grown more slowly (2.5%). Association of Research Libraries. ARL Statistics 2004-05. "Table 4: Expenditure Trends in ARL Libraries, 1986-2005", p. 16.
http://www.arl.org/bm~doc/arlstat05.pdf

[12] Ibid. "Table 3: Supply and Demand in ARL Libraries," p. 14-15. Between 1986 and 2005, serials purchased increased at an average of 1.9% a year—up 42% for the period; monographs purchased decreased at an average of 0.4% a year—down 7% for the period.

[13] OCLC Annual Report 2005/06, p. 14.

[14] Wilder, Stanley J. 2000. The changing profile of research library professional staff. *ARL Bimonthly Report* 208/209.

[15] Wilder, Stanley J. 2002. Demographic trends affecting professional technical services staffing in ARL libraries. *Cataloging and classification quarterly* 34:1/2, p. 53-57.

[16] Wilder, Stanley J. 2007. The new library professional. *Chronicle of higher education* 53:25.

[17] In a forthcoming update for the *ARL Bimonthly Report*, "The ARL youth movement: not what it was," Wilder reports that the ARL population retirement wave will peak in the five year period beginning in 2010. In his earlier publication on professional technical services staffing, he notes the possibility that "fully one-third of the 2000 ARL cataloging population will retire by 2010." E-mail and phone conversations with author, June 13, 2007. Information used with permission.

[18] Based on the presentations at the June 22, 2007 ALCTS preconference "What they don't teach in library school," job descriptions of newly hired librarians to technical services show a trend to blend (or even de-emphasize) traditional cataloging duties with of management, systems, metadata, e-resources, scholarly publishing, and other duties.
http://www.ala.org/ala/alcts/alctsconted/alctsceevents/alctspreconf/LibrarySchool.htm

[19] The figure is inspired by the Mellon-funded research at the University of Minnesota Libraries "A multi-dimensional framework for academic support." The UM's research employed a user-centered analytical methodology based on identifying scholarly information-seeking activities ("primitives") and behavioral trends, followed by the identification of potential new services. For an example of this user-centered approach to library service planning, see
http://www.ala.org/ala/alcts/alctsconted/alctsceevents/alctspreconf/LibrarySchool.htm

[20] O'Reilly, Tim. 2005. What is Web 2.0.
http://oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[21] OCLC Annual Report 2005/06, p. 3.

[22] For descriptions of these programs see
http://www.oclc.org/customcataloging/services/contract/ and
http://www.oclc.org/catalogingpartners/default.htm

[23] ILL Fee Management reduces libraries' administrative costs by tracking and reconciling interlibrary loan charges through the library's OCLC bill. For more information see http://www.oclc.org/resourcesharing/features/feemanagement/default.htm

[24] Information about the Terminologies Service is at http://www.oclc.org/research/projects/termservices/. For as discussion of WorldCat Identities see Lorcan Dempsey's Feb. 13, 2007 weblog at http://orweblog.oclc.org/archives/001269.html

[25] O'Reilly, What is Web 2.0.

[26] For a discussion on Thom Hickey's blog of the new list-making functionality, see http://outgoing.typepad.com/outgoing/2007/06/worldcat-lists.html